Vol. 66, No. 18 Whole No. 350, 1952

Psychological Monographs: General and Applied

# Judging Interests from Expressive Behavior[1]

N. L. Gage
Bureau of Educational Research, College of Education,
University of Illinois

## I. INTRODUCTION

This exploratory study deals with some of the problems involved in measuring the accuracy with which one person perceives another. Such problems are important because behavior in response to stimuli coming from other persons, i.e., social behavior, depends largely on perceptions of other persons. Whether it is a teacher interacting with a pupil, a salesman with a customer, a psychotherapist with a client, or a foreman with a worker, the resulting behavior depends for its effectiveness, however conceived, on the perceptions one person has of the other.

One dimension of such social perceptions is "accuracy," the veridical value, or "truth," in one person's perception of another. It is with certain aspects of this dimension of social perception that we are especially concerned. "Accuracy" is, of course, not the only conceivable dimension of a social perception, nor perhaps the most important for effectiveness in interpersonal relationships. Perhaps such a dimension as "idealization," the degree of perception of others as approximations to one's ideals, is of greater significance; perhaps perceptions of relationships like "dependence," "independence," and "interdependence" will prove ultimately to be of more importance. But in the present research it is the "accuracy" dimension with which we begin.

Secondly, the evidence on which perceptions are based can vary from a casual glance to the profound data elicited by the psychoanalyst. In everyday life, individuals often base their perceptions largely on casual visual and aural data, on brief conversations in a ready-made social context. This investigation approximates these conditions by putting the persons-to-be-perceived before the perceivers only briefly and asking them to provide relatively simple kinds of expressive behavior.

Third, the kinds of perceptions or understandings asked of the perceiver can vary from the superficial to the deep and complex. Thus, the perceiver may be asked to judge merely the age and marital status of a person, or he may be asked to perceive basic motive patterns. The question of what people actually can perceive, or think they can perceive, in various kinds of social stimuli is one that merits investigation in its own right. In this study, the social perceptions asked for were of fair depth; we asked for perceptions of the self-concept of other per-

[1] I am grateful to Dr. Lee J. Cronbach for several imporant suggestions and criticisms and to Mr. Ralph V. Exline for help in collecting data and in preliminary discussions of the results.

This study was supported by research funds from the Bureau of Educational Research, College of Education, University of Illinois.

sons as expressed in answers to an interest inventory.

In short, this exploration of social perceptions deals with a particular selection of possible phenomena in the general area; it is concerned with the *accuracy* of perceptions of the *self-concept* of other persons on the basis of relatively *simple and brief* social stimuli.

Data collected on such phenomena can be used to answer a variety of questions. The following paragraphs epitomize the issues raised in this report.

You see a young man walking down the street. His dress, his gait, his apparent age, and a host of other cues give you an immediate impression of his personality. How accurate is that impression? You are asked to put down your impression by predicting how he would answer a series of questions. Which would he most and least prefer to do: visit an art gallery, browse in a library, or visit a museum? Which course in school would he most and least prefer to take: sketching, biology, or metal working? Your predictions of his answers could be verified by asking the young man these questions and comparing your predictions with his actual responses. Would you be right more often than chance would allow? Would your score over a series of such predictions differ reliably from that of other predictors? Is your performance relatively the same from one series of questions to the next? From one stranger to the next?

Similarly, how is your success determined? Do you actually integrate consciously and unconsciously a host of cues from expressive behavior? Or is your success determined mainly by the accuracy of your stereotype concerning the cultural subgroup to which the stranger belongs?

Furthermore, is your ability to predict strangers' responses related to your intelligence? To your interests in various kinds of activity? To your emotional or social adjustment?

Finally, and most important to us, does ability to judge others have any bearing on effectiveness in interpersonal relationships? On a common-sense level, any layman will agree that people who understand others better should "deal with" others more effectively. Psychologists have operated on the same hypothesis: curricula for training leaders, teachers, foremen, and clinical psychologists are largely aimed at improving sensitivity to and understanding of followers, pupils, employees, and clients. Such improved understanding, it is assumed, will result in more effective leadership, teaching, supervision, and therapy. Consequently many tests have been developed for appraising the knowledges and attitudes that, according to experts, reflect valid understandings of human behavior and its significance. The present enterprise is aimed at developing more direct techniques for appraising insight into other persons. We are interested in an ability to predict others' responses as a possible lead to an important aspect of effectiveness in interpersonal relationships.

We have space for only a few references to the bulky but uncoordinated literature. Allport (1) discussed the general problems involved in studying "the ability to judge other people." Wolff (15) presented much evidence that judgments of personality from brief observations of expressive behavior could be made with significantly better than chance success. Estes (6) showed that strong individual differences in ability to rate others' personalities could be revealed when the ratings were made after viewing brief motion pictures of strangers' expressive behavior. More recently, Dymond (5) has investigated "empathy"; Fiske and his co-workers (7), the diagnostic competence of clinical psychologists; and Bender and Hastorf (3), the perception of persons as reflected in forecasts of friends' responses on three personality scales.

As to the relationship between ability to predict others' responses and effectiveness in interpersonal relationships, previous research has dealt only with the prediction of group responses. Chowdhry (4), Wood (16), and Gage and Suci (8) have reported positive relationships between (*a*) accuracy in predicting the number or percentage of a group who would answer attitude items in a specified way, and (*b*) acceptance by the group predicted. Chowdhry's study established the important qualification that the attitudes predicted must be related to the purposes of the group.

But good predictors of group opinion might have the wrong individuals in mind in making their estimates. Judgments of individuals can be studied more meticulously as to the nature of the processes involved. For this reason, in part, the present study has dealt with predictions of individual behavior. Similarly, we wished to determine whether the ability to predict can function on the basis of relatively brief acquaintance with strangers, rather than of longer acquaintanceships with fellow members of an organization, classmates, or pupils. If such a prediction-from-brief-acquaintance ability exists, does it have some functional significance for effectiveness in interpersonal relationships? If it does, we shall have the possibility of developing "transportable" tests, nonspecific to a given set of interpersonal relationships. Such tests would be useful in the selection and training of persons whose work requires them to achieve quick and accurate understandings of other people.

Our report deals with a novel combination of two approaches used in previous studies. First, we shall be concerned with what Allport called *explicit forecasts* (1, p. 385) of specific behaviors, i.e., responses to test items; this is the form of behavior-to-be-predicted that has previously been used by Bender (2) and Steinmetz (12). Secondly, we shall be dealing with predictions of *strangers'* behavior, i.e., predictions made after brief observations of the expressive behavior of persons otherwise unknown; this is the kind of evidence for prediction that has been supplied to judges in the work of Wolff (15) and Estes (6).

We must, of course, be concerned first with whether an ability to make explicit forecasts of others' responses actually exists as a trait in the sense of generality over strangers, items, and trials. If this question can be answered affirmatively, we are justified in seeking for the correlates of this ability. In this paper we report evidence on this question and on the meaning of the ability tested, in terms of the relationship of the measures to sociometric status and other characteristics of the judges.

## II. PROCEDURE

Four trials of the following exercise have been carried out: six strangers, three male alternating with three female, were brought one at a time before a roomful of judges. The strangers were previously unknown to the judges. Each stranger was instructed to go through the following set of tasks, patterned after Estes' (6), designed to elicit expressive behavior: (*a*) Go to the blackboard and draw a house, a tree, and a person; (*b*) tear a match from a book of matches, light it, and hold it as long as possible; (*c*) using a deck of playing cards, build on a table the most elaborate house he can in two minutes; (*d*) play a game of crisscross slap hands with an assistant for about a minute; and (*e*) describe the room he is in, in any way he likes, taking about two minutes.

Upon completing the last task, the stranger left the room and filled out the first two pages of the Kuder Preference Record, Form BM. (In the third and fourth trials, the first three pages of the Kuder were used.) At the same time, the judges answered the Kuder *as they predicted the stranger would.* The judge was scored by using the stranger's answer sheet as the scoring key.

The scoring was done so as to weight the prediction for each Kuder item according to the number of correct discriminations made. Each Kuder item consists of three activities from which the stranger is asked to choose the one he would most like to do and the one he would least prefer. If the three activities are labeled A, B, and C, and the stranger chooses A as most and C as least preferred, then the stranger's order of preference is A-B-C. This pattern would be considered the "correct" one for this item in scoring the judge's prediction. Now consider

how each of the judge's possible predictions is scored:

| Possible Prediction (Correct Pattern A-B-C) | Number of Correct Discriminations | Score (No. of Points Given) |
|---|---|---|
| A-B-C | 3 | 2 |
| A-C-B | 2 | 1 |
| B-A-C | 2 | 1 |
| B-C-A | 1 | 0 |
| C-A-B | 1 | 0 |
| C-B-A | 0 | −1 |

Each score, it will be noted, is one less than the number of correct discriminations made by the given possible prediction. This arrangement maximized ease of scoring.

A window stencil was made which had holes punched through the answer spaces marked on the answer sheet by the stranger. When it was placed over the judge's prediction answer sheet, the number of marks seen through the window stencil provided the judge's score except in the case of zero correct discriminations. To detect these, another window stencil was made with the holes punched in exactly the opposite of the correct pattern for each item. Whenever *both* of the judge's marks for a given item could be seen through this "reverse," or wrong, stencil, he was given a score of −1 for that item.[2]

Training Club at the University of Illinois. The six strangers were students in an introductory course in educational psychology. In addition to judging strangers, the judges took the Strong Vocational Interest Blank for Men, the Kelley-Perkins How I Teach inventory (9), and the Cook-Leeds Teacher Attitudes Inventory (10); for 12 of the judges we also had scores on the ACE Psychological Examination for College Freshmen. The two inventories of teachers' attitudes were used to provide measures of understanding of others by a different approach from that involved in our task of judging strangers.

*Generality over items.*—If judges predict strangers' responses to one set of Kuder items, will they do so with the same relative success on an equivalent set of Kuder items? This is the traditional split-half, or equivalence, conception of reliability. It was applied here by scoring each judge for his success in predicting each stranger's responses on the odd-numbered and, separately, on the even-numbered items of pages one and two of the Kuder. When these scores were correlated for each stranger separately, and for sums over all six strangers, we obtained the $r$'s shown in Table 1. When corrected with the Spearman-Brown formula, the $r$'s for single stran-

TABLE 1

TRIAL ONE

GENERALITY OVER ITEMS: CORRELATIONS BETWEEN ACCURACY SCORES ON ODD- AND EVEN-NUMBERED ITEMS FOR EACH STRANGER AND ALL SIX STRANGERS

$N = 24$

| Score | Stranger | | | | | | Total |
|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | |
| Mean | 14 | 19 | 14 | 13 | 20 | 18 | 98 |
| *SD* | 5 | 6 | 4 | 4 | 4 | 3 | 15 |
| $r_{uncorrected}$ | .12 | .41 | .35 | .50 | .33 | .02 | .65 |
| $r_{corrected}$ | .21 | .58 | .52 | .67 | .50 | .04 | .79 |

Since the four trials differed in the kinds of strangers, judges, and additional data involved, they have been analyzed and are here reported separately.

## III. TRIAL ONE

In this trial the judges were 24 members of an undergraduate Teachers-in-

[2] This scoring method resembles the two suggested by Harold Gulliksen (*Theory of Mental Tests*. New York: Wiley 1950, p. 259). It is, however, less arbitrary in its weighting and provides the full range of scores.

gers range from .04 to .67 with a median of about .50. The corrected odd-even $r$ for all six strangers is .79. It appears that the ability here measured has substantial generality over items, although for strangers A and F individually this statement does not hold.

*Generality over strangers.*—If judges predict one stranger's responses, will they do so with the same relative success for another stranger? In this conception of generality, the strangers are analogous to the tests of the traditional equivalent-forms method of estimating reliability. For our data this means obtaining the intercorrelations of the scores on the six strangers individually, and on the strangers grouped by threes. Table 2 shows the $r$'s obtained. All but two of the 15 $r$'s are positive, the range being from −.24 to .40, with the median at .21. These $r$'s indicate fairly consistently that the answer to our question is affirmative. We can estimate the generality over all six strangers by using an analogue of the Kuder-Richardson formula 20:[3]

TABLE 2

TRIAL ONE

GENERALITY OVER STRANGERS: INTERCORRELATIONS AMONG ACCURACY SCORES ON SIX STRANGERS AND BETWEEN SCORES ON FIRST AND SECOND THREE STRANGERS

$N = 24$

| | Stranger | | | | | |
|---|---|---|---|---|---|---|
| | A | B | C | D | E | F |
| A | | .22 | .08 | .16 | .28 | .24 |
| B | | | .22 | .40 | .21 | −.24 |
| C | | | | .14 | .24 | .21 |
| D | | | | | .21 | .04 |
| E | | | | | | −.10 |

$r_{(A+B+C)(D+E+F)} = .42$, correcting to .59.
Median $r = .21$.
Kuder-Richardson 20 = .53.

$$r = \frac{n}{n-1}\,\frac{2\Sigma\sigma_i\sigma_j r_{ij}}{\Sigma\sigma_i^2 + 2\Sigma\sigma_i\sigma_j r_{ij}}$$

where $n$ = number of strangers
$\sigma_i$ = standard deviation of prediction scores on stranger $i$
$\sigma_j$ = standard deviation of prediction scores on stranger $j$
$r_{ij}$ = correlation between prediction scores for strangers $i$ and $j$

Applied here, this formula yields an $r$ of .53.

When the scores on the first three strangers are correlated with those on the second three, the $r$ is .42. Correcting for a "test" twice as long (i.e., for six strangers), we obtain an $r$ of .59.

By both approaches, thus, we can estimate the generality of the ability to judge all six strangers in terms of a coefficient of about .55.

*Generality over strangers and items.*—If judges predict the responses of one group of strangers to one set of items, will they do so with the same relative success for another group of strangers' responses to another set of items? In this conception of generality, we have the most rigorous test of whether we have actually identified a common trait. For our data this means obtaining the correlation between the judges' scores on the odd-numbered items for the first three strangers and their scores on the even-numbered items for the second three strangers. Similarly, we need the correlation between the scores on the even-numbered items for the first three strangers and the scores on the odd-numbered items for the second three strangers. Both these $r$'s, along with other estimates of reliability over items and over strangers, are shown in Table 3. The two $r$'s relevant here (starred in Table 3) are .31 and .48.

[3] Use of this formula was suggested by L. J. Cronbach.

TABLE 3

TRIAL ONE

GENERALITY OVER ITEMS AND STRANGERS: CORRELATIONS BETWEEN ACCURACY SCORES ON ODD-NUMBERED ITEMS FOR FIRST THREE STRANGERS AND SCORES ON EVEN-NUMBERED ITEMS FOR SECOND THREE STRANGERS, ETC.

$N = 24$

| | 1st 3 Odd | 1st 3 Even | 2nd 3 Odd | 2nd 3 Even |
|---|---|---|---|---|
| 1st 3, Odd | | .41 | −.01 | .48* |
| 1st 3, Even | | | .31* | .35 |
| 2nd 3, Odd | | | | .34 |

* These *r*'s indicate reliability over items *and* strangers.

When these are corrected to yield estimates for a "test" four times as long (i.e., one with twice as many strangers and items), we obtain *r*'s of .64 and .79. Averaging roughly, we may express the generality over items and strangers in the present task, involving six strangers and 20 Kuder items, at about .70.

*Correlations with other measures.*—Having established the adequate generality of our measures of ability to judge strangers' interests from their expressive behavior, we may explore the meaning of the scores in terms of their correlations with other measures. The *r*'s we obtained are shown in Table 4. Since an *r* has to equal .41 to be significant at the 5 per cent level (when *N* equals 23), it is evident that none of our obtained coefficients is significantly different from zero.

Our essentially zero correlations with the Strong scores for Artists and Musicians are especially interesting in the light of the positive relationships with aesthetic interests reported by Vernon (13) and Estes (6). The latter's conclusion is especially called in question, since he worked with ability to judge strangers on the basis of brief motion pictures of their expressive behavior, a task similar to the one we set. It may be that the aspects of personality judged in Estes' study (Murray's personality variables) differ in such a way from the responses to Kuder items which we used as to account for the difference. It is also possible, of course, that the more impressionistic evidence concerning their judges' interests may account for the contrary conclusions drawn by Vernon and Estes.

TABLE 4

TRIAL ONE

CORRELATIONS BETWEEN OTHER MEASURES AND THE ACCURACY SCORES

$N = 23$*

| Test Score | Judging Strangers | Strong Vocational Interest Blank for Men | Judging Strangers |
|---|---|---|---|
| How I Teach, Form A | −.09 | Social Studies Teacher | .28 |
| How I Teach, Form B | .04 | Masculinity | .25 |
| How I Teach, Forms A+B | .02 | Sales Manager | .09 |
| Teacher-Attitude Inventory | −.08 | Production Manager | .08 |
| ACE Psychol. Exam., Total ($N = 12$) | −.39 | Ability to Fake Masculinity | −.02 |
| | | President of Manufacturing Concern | −.03 |
| | | Math.-Science Teacher | −.07 |
| | | Psychologist | −.10 |
| | | Chemist | −.15 |
| | | Author-Journalist | −.18 |
| | | Artist | −.28 |
| | | Musician | −.31 |

* For $N = 23$, *r* must equal .41 to be significant at the 5% level.

TABLE 5

TRIAL TWO

GENERALITY OVER ITEMS: CORRELATIONS BETWEEN ACCURACY SCORES ON ODD- AND EVEN-NUMBERED ITEMS FOR EACH STRANGER AND ALL SIX STRANGERS

$N = 18$

| Score | Stranger | | | | | | Total |
|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | |
| Mean | 13 | 24 | 22 | 20 | 11 | 18 | 108 |
| *SD* | 4 | 3 | 4 | 3 | 3 | 3 | 10 |
| $r_{uncorrected}$ | .59 | .29 | .36 | −.31 | .66 | −.11 | .36 |
| $r_{corrected}$ | .74 | .45 | .53 | — | .80 | — | .53 |

## IV. TRIAL TWO

In this trial the judges were 18 in number—6 repeaters from the Teachers-in-Training Club and 12 volunteers from undergraduate educational psychology classes. The strangers were again students from an educational psychology class, with the exception of 2 female strangers who were clerical employees in the university's testing office and wives of students in the university. These 2 strangers, D and F, were thus not members of the same "cultural subgroup" as the other strangers, a fact which may help us understand some of the results.

*Generality over items.*—As is shown in Table 5, the corrected odd-even reliability estimates for four of the strangers were positive, ranging from .45 to .80. But for strangers D and F, the two female clerical employees, the $r$'s between odd and even scores were negative. These strangers were no more "difficult" to judge than the other four, since the mean scores of the judges for these two were exactly in the middle of the range of means. Nor do the low standard deviations of the scores on these two strangers account for the lack of reliability of the judges' predictions of their responses, since the next lowest standard deviation is accompanied by the highest of all the four positive odd-even $r$'s. Unless it is mere coincidence, the explanation for these two negative $r$'s must be sought in the fact that they were not members of the same cultural subgroup (undergraduates in education) to which the judges correctly assigned the other four strangers.

It should be noted that in spite of the negative values for strangers D and F, the generality over items for the other four strangers was sufficient to produce a positive coefficient between total odd and even scores equal to .53 (corrected).

*Generality over strangers.*—Table 6 shows the intercorrelations among the scores on the six strangers. The range is from −.29 to .36, the median falling at −.02. Our analogue of the Kuder-Richardson formula yields $r$ equal to .15.

TABLE 6

TRIAL TWO

GENERALITY OVER STRANGERS: INTERCORRELATIONS AMONG ACCURACY SCORES ON SIX STRANGERS AND BETWEEN SCORES ON FIRST AND SECOND THREE STRANGERS

$N = 18$

| | Stranger | | | | | |
|---|---|---|---|---|---|---|
| | A | B | C | D | E | F |
| A | | .36 | −.09 | −.29 | −.07 | .05 |
| B | | | −.03 | −.09 | .19 | .28 |
| C | | | | −.02 | −.17 | .29 |
| D | | | | | .06 | −.14 |
| E | | | | | | .15 |

$r_{(A+B+C)\ vs.\ (D+E+F)} = -.004$.
Median $r = -.02$.
Kuder-Richardson $20 = .15$.

The correlation between scores on the first three and second three strangers is −.004. Both these approaches indicate that in this trial there was complete absence of generality over strangers in the ability to predict responses to the Kuder Preference Record.

*Generality over items and strangers.*—Table 7 shows the two *r*'s, .08 and −.02,

TABLE 7

TRIAL TWO

GENERALITY OVER ITEMS AND STRANGERS: CORRELATIONS BETWEEN ACCURACY SCORES ON ODD-NUMBERED ITEMS FOR FIRST THREE STRANGERS AND SCORES ON EVEN-NUMBERED ITEMS FOR SECOND THREE STRANGERS, ETC.

$N = 18$

| | 1st 3 Odd | 1st 3 Even | 2nd 3 Odd | 2nd 3 Even |
|---|---|---|---|---|
| 1st 3, Odd | | .50 | .02 | −.02* |
| 1st 3, Even | | | .08* | −.03 |
| 2nd 3, Odd | | | | .20 |

* These *r*'s indicate reliability over items *and* strangers.

between scores that are based on both different items and different strangers. The *r*'s between scores that have items in common are .50 and .20, indicating again the generality over items described above. But the *r*'s between scores based on different strangers are .02 and −.03, showing again the failure to obtain generality over strangers in this trial.

These results were puzzling after the promise of the data from Trial One. To establish more definitely whether the task we had set our judges elicited measures of a common trait, we carried out a third trial.

## V. TRIAL THREE

In this trial the judges were 30 students in an undergraduate educational psychology course, and the strangers were undergraduates in teacher-training curricula hired through the university's student employment office. The first three pages of the Kuder were used rather than two as in the preceding trials.

In addition to the basic task of predicting strangers' responses, the judges were asked *beforehand* to fill out two answer sheets for the first three pages of the Kuder as follows: the first with the responses which the judge thought the *typical male undergraduate in teacher training at the University of Illinois* would make; the second with the responses which the judge thought would be made by the *typical female undergraduate in teacher training at the University of Illinois.* These are subsequently referred to as the "male stereotype" and the "female stereotype" sets of answers, respectively.

Finally, after all strangers had been judged, the judges filled out a complete answer sheet with *their own responses* to all the Kuder items. The judge's scores on the entire Kuder, obtained

TABLE 8

TRIAL THREE

GENERALITY OVER ITEMS: CORRELATIONS BETWEEN ACCURACY SCORES ON ODD- AND EVEN-NUMBERED ITEMS FOR EACH STRANGER AND ALL SIX STRANGERS

$N = 30$

| Score | Stranger | | | | | | Total |
|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | |
| Mean | 26 | 24 | 13 | 22 | 31 | 26 | 142 |
| *SD* | 8 | 6 | 6 | 6 | 5 | 7 | 25 |
| $r_{uncorrected}$ | .64 | .21 | .20 | .45 | .44 | .14 | .60 |
| $r_{corrected}$ | .78 | .35 | .33 | .62 | .61 | .25 | .75 |

with Kuder's scoring keys, are used as measures of the judge's own interests, to be correlated with his success in judging the strangers.

*Generality over items.*—Table 8 shows that the correlations between scores based on odd- and even-numbered items were positive for all six strangers, the corrected $r$'s ranging from .25 to .78. The $r$ between total odd scores and total even scores is .75 (corrected), again indicating that the ability tapped by the judging task is fairly general over the different samples of Kuder items.

TABLE 9

TRIAL THREE

GENERALITY OVER STRANGERS: INTERCORRELATIONS AMONG ACCURACY SCORES ON SIX STRANGERS AND BETWEEN SCORES ON FIRST AND SECOND THREE STRANGERS

$N=30$

| | Stranger | | | | | |
|---|---|---|---|---|---|---|
| | A | B | C | D | E | F |
| A | | .13 | .46 | .15 | .45 | .49 |
| B | | | .03 | .23 | −.13 | −.15 |
| C | | | | .36 | .42 | .13 |
| D | | | | | .04 | .24 |
| E | | | | | | .23 |

$r_{(A+B+C)(D+E+F)}=.55$, correcting to .71.
Median $r=.23$.
Kuder-Richardson 20 = .61.

*Generality over strangers.*—Table 9 shows that the intercorrelations among the scores on the six strangers range from −.15 to .49, the median $r$ equalling .23. The adapted Kuder-Richardson formula yields an estimate of .61 for the generality over strangers of our task. The $r$ between scores on the first and second three strangers is .55, which corrects to .71 for a "test" twice as long. These results corroborate those of Trial One in indicating a substantial degree of generality of the ability tested from one stranger to the next.

*Generality over items and strangers.*—

TABLE 10

TRIAL THREE

GENERALITY OVER ITEMS AND STRANGERS: CORRELATIONS BETWEEN ACCURACY SCORES ON ODD-NUMBERED ITEMS FOR FIRST THREE STRANGERS AND SCORES ON EVEN-NUMBERED ITEMS FOR SECOND THREE STRANGERS, ETC.

$N=30$

| | 1st 3 Odd | 1st 3 Even | 2nd 3 Odd | 2nd 3 Even |
|---|---|---|---|---|
| 1st 3, Odd | | .48 | .36 | .37* |
| 1st 3, Even | | | .49* | .44 |
| 2nd 3, Odd | | | | .29 |

* These $r$'s indicate reliability over items *and* strangers.

In Table 10, the two starred $r$'s, .37 and .49, show that there is considerable generality over items and strangers of the ability measured in this trial. Corrected for a "test" four times as long, these estimates become .70 and .79.

All in all, the results of Trial Three are in striking agreement with those of Trial One. Whatever was being measured in these trials appears to be a trait that can be measured with substantial consistency. The variance between judges is distinctly nonrandom, exhibiting definite generality over items and strangers.

*Analysis of the judging process.*—With this trial we were able to undertake an analysis of the nature of the ability being tapped by the task of predicting strangers' responses. It will be remembered that we had four sets of answers in this trial:

1. Stranger. The stranger's own responses to the first three pages of the Kuder.
2. Prediction. The judge's predictions of the stranger's responses; there were six such answer sheets for each judge.
3. Stereotype. The judge's predictions of the responses of the typical male, and typical female, teacher-in-training at the University of Illinois; these answer sheets were, as noted above, filled out before any of the strangers was seen.
4. Judge. The judge's own answers given according to how he himself felt about each of the items of the Kuder Preference Record.

When the predictions are matched against the stranger's responses we obtain, as already described, the *accuracy* scores. When the stereotypes are matched against the stranger's responses, we can obtain what we shall term *stereotype* scores, that is, the accuracy with which the stranger's responses would have been predicted from the stereotype alone, without the stranger's ever having been observed.

It is noteworthy that the mean of the stereotype scores, 160, is higher than that of the prediction scores, 142. This difference is significant at the 1 per cent level. Both means are significantly higher than mathematical chance, which would yield a score of 90, but the difference between them is indeed opposite in direction from common-sense expectation. That is, it seemed reasonable to expect that the predictions from a knowledge merely of the cultural subgroup to which the strangers belonged would be *less* accurate than those made not only after this knowledge but also after observations of the strangers' expressive behavior. But these results suggest that the clues to personality, which the expressive behavior should have provided, either did not appear or were misused by the judges.

This finding suggests that the process of judging strangers is much different from what we had supposed. Judges do not integrate a host of subtle clues from expressive behavior in arriving at judgments concerning the strangers. Rather they make inferences from their relatively gross knowledge of the subgroup to which they think the stranger belongs and, insofar as the observation of expressive behavior affects their judgments at all, the effect is to reduce accuracy.

Our results differ from those of Smith (11), who compared the mean scores of judges in answering factual questions about the early years of a young man before and after reading his detailed cross-sectional, autobiographical case study. His 18 judges, graduate students in a seminar on the life-history method, made scores slightly poorer than mathematical chance before reading the case study; after reading it, their scores were significantly better than both the psychological and mathematical zero points. The difference in our results probably stems from the facts that Smith initially gave his judges no information about the stranger other than that to be inferred from the questions themselves, while we gave the judges the cultural subgroup (male or female teachers-in-training at the University of Illinois) to which the strangers belonged. Further, it is possible that Smith's case-study material was more relevant to the predictions made than was our strangers' expressive behavior. Hence Smith's results and ours are comparable only in the general nature of the problem confronted, but not as to procedure and results.

A further method of ascertaining the meaning of the predictive-accuracy scores is to compare them with the scores obtained when each prediction is matched against (*a*) the responses of the other two strangers of the same sex (accidental like-sex accuracy), and (*b*) the responses of the other three strangers of the opposite sex (accidental cross-sex accuracy). If the predictions are made with no discrimination beyond that possible from knowing the sex of the stranger, the mean of the "accidental like-sex accuracy" scores should be as great as the predictive-accuracy mean. Similarly, if the predictions did not even benefit from being aimed at strangers of a given sex, then the accidental cross-sex accuracy scores should be as high as the accidental like-sex accuracy scores. It is also possible to compare the stereotype accuracy with the scores obtained when the stereotype predictions are matched against the three strangers of the sex opposite to that for which the stereotype was intended.

Table 11 shows the means of these five kinds of scores for each stranger and for all strangers. It is apparent that, for five

TABLE 11

TRIAL THREE

MEAN SCORES FOR PREDICTIVE ACCURACY, ACCIDENTAL LIKE-SEX ACCURACY, ACCIDENTAL CROSS-SEX ACCURACY, STEREOTYPE ACCURACY, AND STEREOTYPE CROSS-SEX ACCURACY

$N = 26$

| Score | Stranger Predicted | | | | | | Total |
|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | |
| Predictive Accuracy[1] | 26 | 25 | 14 | 22 | 30 | 26 | 143 |
| Accidental Like-Sex Accuracy[2] | 23 | 23 | 15 | 19 | 25 | 22 | 127 |
| Accidental Cross-Sex Accuracy[3] | 21 | 20 | 15 | 14 | 21 | 25 | 116 |
| Stereotype Accuracy[4] | 30 | 27 | 23 | 28 | 28 | 24 | 160 |
| Stereotype Cross-Sex Accuracy[5] | 26 | 20 | 24 | 22 | 16 | 31 | 140 |

[1] Predictions of individual strangers scored against the responses of the strangers for whom the predictions were intended.

[2] Predictions of individual strangers scored against the responses of the other two strangers of the same sex.

[3] Predictions of individual strangers scored against the responses of the other three strangers of the opposite sex.

[4] Predictions of the responses of the "typical" person of a defined sex and curricular group scored against the responses of the three strangers of that sex.

[5] Predictions of the responses of the "typical" person of a defined sex and curricular group scored against the responses of the three strangers of the opposite sex.

of the six strangers, the predictive accuracy was greater than the accidental like-sex accuracy. Thus the predictions fit the stranger for whom they were intended better than they fit the other two strangers of the same sex. In other words, the predictions benefited from being aimed at the specific stranger. The total accuracy for all six strangers, 142, is greater than the total accidental like-sex accuracy, 127.

Further, for four of the six sets of predictions, the accidental like-sex accuracy is greater than the accidental cross-sex accuracy; the total of the former scores over the six strangers, 127, is greater than that of the latter, 116. Thus the predictions tended to fit other strangers of the same sex as the stranger for whom they were intended better than they fit other strangers of the opposite sex. In short, the predictions benefited from being aimed at strangers of a given sex.

The stereotype predictions also fit strangers of the intended sex better than they fit those of the opposite sex except for strangers C and F, who were also the exceptions to the trend noted in the preceding paragraph. Over all six strangers the total stereotype accuracy, 160, was higher than the total stereotype cross-sex accuracy, 140. Here again the stereotype benefited from a knowledge of the sex of the stranger.

The data of Table 11 may be summarized as indicating that, while predictions of individuals did not agree with the responses of individuals as closely as did the "curriculum-sex" stereotype, they did, nonetheless, fit the individual better than they fit his fellows of the same sex and still better than they fit the strangers of the opposite sex. In short, the predictions show signs of some success on the part of the judges in fitting their predictions to a specific stranger of a given sex.

*Correlations of accuracy scores with judges' own interests.*—Regardless of the processes underlying the scores, it is of some potential interest to ascertain the relationship between these scores and

the judges' own interests as measured by Form BM of the Kuder Preference Record. We should expect the most positive of the relationships to be that with "Social Service" interests, since this score reflects the kind of values and activity preferences, say "interest in people," which should lead to the development of ability to judge strangers. As is shown in Table 12, this expectation is borne out, the $r$ of .21 with the social-service score being the highest positive one obtained. All $r$'s are so low, however, and the differences between them so small, that this finding is easily ascribable to chance.

TABLE 12

TRIAL THREE

CORRELATIONS BETWEEN JUDGES' SCORES ON THE KUDER PREFERENCE RECORD, FORM BM, AND THEIR ACCURACY SCORES

$N=30$

| Kuder Score | Mean | *SD* | $r$ with Accuracy Score |
|---|---|---|---|
| Mechanical | 53 | 19 | −.07 |
| Computational | 22 | 9 | .09 |
| Scientific | 52 | 10 | .06 |
| Persuasive | 73 | 13 | .12 |
| Artistic | 57 | 14 | −.02 |
| Literary | 48 | 12 | −.13 |
| Musical | 25 | 8 | −.14 |
| Social Service | 93 | 16 | .21 |
| Clerical | 44 | 14 | −.10 |

## VI. TRIAL FOUR

In this trial the judges were the forty-odd members of the senior class at the University High School of the University of Illinois. The six strangers were University of Illinois freshmen hired through the student employment office. To increase the judges' motivation beyond the apparently substantial level achieved by explaining to them the general purpose of the experiment, they were offered \$6, \$3, and \$1 prizes for the best, second-best and third-best overall prediction scores.

The following data were collected in the order indicated:

First, the judges filled out answer sheets with their own answers to the first three pages of the Kuder Preference Record.

Second, the judges answered the following six questions concerning their classmates:

1. What three boys and three girls in your class do you consider to be among your closest friends?
2. Assume that everyone in your class is equally available and willing to do schoolwork with you. With what three boys and three girls would you most like to do your schoolwork?
3. With what three boys and three girls in your class would you most like to be when you are at social activities, such as school dances and parties?
4. What three boys and three girls in your class are best at catching on quickly to how other students feel about things?
5. What three boys and three girls in your class are best at being tactful, that is, at not hurting others' feelings?
6. What three boys and three girls in your class do you think would be quickest and most accurate at "sizing up" a newcomer to the class?

These questions were intended to provide a measure of the acceptance of the students by their classmates and of their effectiveness in interpersonal relations as judged by their classmates.

Third, the judges answered the first three pages of the Kuder as they thought the typical University High School male senior would and then as they thought the typical University High School female senior would. These provided the "stereotype" answer sheets as in Trial Three.

Fourth, the judges observed each of the six strangers in the expressive behavior already described and then filled out an answer sheet for the first three pages of the Kuder as they thought the stranger would.

Fifth, the judges filled out similar

TABLE 13

TRIAL FOUR

GENERALITY OVER ITEMS: CORRELATIONS BETWEEN ACCURACY SCORES ON ODD- AND EVEN-NUMBERED ITEMS FOR EACH STRANGER AND FOR ALL STRANGERS

$N=40$

| Score | Stranger | | | | | | Total |
|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | |
| Mean | 20 | 24 | 20 | 29 | 26 | 25 | 144 |
| *SD* | 6 | 5 | 6 | 6 | 6 | 7 | 21 |
| $r_{uncorrected}$ | .25 | .29 | .16 | −.04 | .29 | .30 | .54 |
| $r_{corrected}$ | .40 | .45 | .28 | — | .45 | .46 | .70 |

answer sheets with answers they thought would be made by each of three of their classmates selected by chance.

In addition to these data it was possible to secure from the school's guidance office the scores of a majority of the seniors on the Henmon-Nelson Test of Mental Ability, the ACE Psychological Examination for High School Students, the Bell Adjustment Inventory, and the Kuder Preference Record, Form CH.

*Generality.*—Tables 13, 14, and 15 show the results of the analysis to estimate generality carried out in the same way as those for the first three trials. In general, the results for Trial Four agree with those from Trials One and Three in yielding estimates of generality over items of about .70, over strangers of about .60, and over items and strangers of about .60.

In this trial, we had the judges' scores in predicting not only the strangers' responses to the Kuder Preference Record but also those of three of their classmates selected at random. How does accuracy in predicting strangers' interests correlate with accuracy in predicting classmates'? The $r$, for the 34 judges who furnished both sets of scores, is .71. From this we can infer that the processes involved in the two kinds of tasks are highly similar. Whether the accuracy scores reflect accuracy in using clues from expressive behavior or some other process, the judges had the same relative success in predicting interests of classmates and of strangers. Pending corroboration

TABLE 14

TRIAL FOUR

GENERALITY OVER STRANGERS: INTERCORRELATIONS AMONG ACCURACY SCORES ON SIX STRANGERS AND BETWEEN SCORES ON FIRST AND SECOND THREE STRANGERS

$N=40$

| | Stranger | | | | | |
|---|---|---|---|---|---|---|
| | A | B | C | D | E | F |
| A | | .04 | .25 | .14 | .21 | .04 |
| B | | | .10 | .06 | −.01 | .20 |
| C | | | | .25 | .38 | .27 |
| D | | | | | .25 | .20 |
| E | | | | | | .41 |

$r_{(A+B+C)(D+E+F)}=.40$, correcting to .57.
Median $r=.20$.
Kuder-Richardson 20 = .59.

TABLE 15

TRIAL FOUR

GENERALITY OVER ITEMS AND STRANGERS: CORRELATIONS BETWEEN ACCURACY SCORES ON ODD-NUMBERED ITEMS FOR FIRST THREE STRANGERS AND SCORES ON EVEN-NUMBERED ITEMS FOR SECOND THREE STRANGERS, ETC.

$N=40$

| | 1st 3 Odd | 1st 3 Even | 2nd 3 Odd | 2nd 3 Even |
|---|---|---|---|---|
| 1st 3, Odd | | .47 | .38 | .16* |
| 1st 3, Even | | | .39* | .26 |
| 2nd 3, Odd | | | | .43 |

* These $r$'s indicate reliability over items *and* strangers.

of this finding, we can say that the task of predicting strangers' interests has much in common with that of judging the interests of acquaintances of fairly long standing.

With this fourth trial, we may conclude that the findings concerning generality are well established. Predictions of the responses of strangers on the first three pages of the Kuder Preference Record are correct or incorrect to a degree which is sufficiently consistent from item to item, or from stranger to stranger, to justify the conclusion that we are dealing with a measurable ability. The evidence supports the idea of a common trait, a recurring tendency for our judges to have the same relative degree of success in predicting strangers' responses from one item to the next, or from one stranger to the next.

TABLE 16

TRIAL FOUR

MEAN SCORES FOR PREDICTIVE ACCURACY, ACCIDENTAL LIKE-SEX ACCURACY, ACCIDENTAL CROSS-SEX ACCURACY, STEREOTYPE ACCURACY, AND STEREOTYPE CROSS-SEX ACCURACY

$N=26$

| Score* | Stranger Predicted | | | | | | Total |
|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | |
| Predictive Accuracy | 20 | 24 | 20 | 29 | 26 | 25 | 144 |
| Accidental Like-Sex Accuracy | 21 | 22 | 19 | 22 | 26 | 22 | 132 |
| Accidental Cross-Sex Accuracy | 15 | 10 | 18 | 14 | 11 | 11 | 79 |
| Stereotype Accuracy | 23 | 27 | 24 | 25 | 25 | 29 | 153 |
| Stereotype Cross-Sex Accuracy | 13 | 16 | 16 | 5 | 22 | 14 | 86 |

* For definition, refer to footnotes of Table 11.

*Analysis of the judging process.*—In this trial, we had the same four kinds of answer sheets as in Trial Three.

Table 16 shows that the mean of the stereotype scores (153) is again higher than that of the accuracy scores (144); this difference is significant at the 3 per cent level. Again it signifies that judges would have done better if they had been uninfluenced by whatever clues to the stranger's personality they picked up from their observation of expressive behavior. This higher stereotype mean score is all the more challenging to our common-sense notions as to how the judging process operates when we remember how the stereotype answer sheets were filled out. The judges marked these sheets for typical University High School seniors rather than for typical University of Illinois freshmen, which was the cultural subgroup to which the strangers actually belonged. Nonetheless their mean stereotype score surpassed their mean accuracy score.

Table 16 also shows that the mean scores for accuracy, accidental like-sex accuracy, and accidental cross-sex accuracy differ similarly as in Trial Three. Again, although the predictions for individuals were not as accurate as the stereotypes when scored against the actual responses of individual strangers, they show distinct evidence of having been influenced toward greater accuracy by being aimed at specific individuals of a given sex.

*Correlations with sociometric status.*—Another paper (8) has presented the rationale by which we hypothesize that accuracy of social perception is positively

related to effectiveness in interpersonal relationships. Our reasoning is that accuracy of social perception, either as a link between certain value systems and ways of behaving in relationship to other persons, or as a consequent of the value systems on the same level as ways of behaving, should be found to vary concomitantly with measures of effectiveness in interpersonal relations.

In the present trial, our measures of effectiveness in interpersonal relationships were obtained by counting the number of mentions each student received from his classmates in answer to the six questions listed above. Number of mentions was taken as a rough approximation to the standing of each University High School senior in his class. Correlations of the six "number of mentions" scores with the accuracy scores are given in Table 17.

TABLE 17

TRIAL FOUR

RELATIONSHIPS OF VARIOUS SCORES TO SOCIOMETRIC INDICES

$N=40$

| Score | Sociometric Indices | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 Closest Friends | 2 Doing School-work | 3 Social Activi-ties | 4 Catching on to Students' Feelings | 5 Being Tactful | 6 Sizing up New-comer | Total |
| Accuracy | .16 | .35 | .24 | .37 | .31 | .40 | .37 |
| Stereotype | .27 | .35 | .24 | .36 | .41 | .38 | .41 |
| Similarity | .31 | .07 | .34 | .37 | .29 | .34 | .34 |
| Rigidity | .27 | .29 | .41 | .34 | .26 | .41 | .38 |
| Assumed Similarity | .29 | .14 | .36 | .38 | .24 | .33 | .33 |
| Self-Typicality | .31 | −.09 | .21 | .07 | −.006 | .08 | .10 |
| Stereotype+Rigidity | .31 | .34 | .42 | .42 | .35 | .48 | .45 |
| Similarity+Rigidity | .32 | .26 | .45 | .41 | .31 | .46 | .42 |
| Accuracy−Stereotype | −.06 | .06 | .08 | .03 | −.06 | .07 | .03 |
| Assumed Similarity−Similarity | .17 | .11 | .25 | .25 | .12 | .20 | .21 |
| Henmon-Nelson IQ ($N=24$) | .19 | .44 | .20 | .16 | .30 | .25 | .27 |
| ACE Psychological Examination ($N=26$) | .26 | .60 | .23 | .23 | .41 | .29 | .39 |

On a priori grounds, it should be expected that the number of mentions for Question 6, "What three boys and three girls in your class do you think would be quickest and most accurate at 'sizing up' a newcomer to the class?" should be most highly related to scores for ability to predict responses of strangers. This is what we find, the $r$ in this case (.40) being the highest of the six. The second highest $r$, .37, is that with a highly similar question, "What three boys and three girls in your class are best at catching on quickly to how other students feel about things?" These $r$'s support the validity of our accuracy scores, using as the criterion the classmates' judgments as to the ability of classmates along lines which a priori seem highly related to what our accuracy scores measure.

The correlation with Question 5, "What three boys and three girls in your class are best at being tactful, that is, at not hurting others' feelings?" does get more directly at effectiveness in interpersonal relationships. Similarly, we may take the kind of interpersonal relationships involved in collaboration on school work, Question Two, "With what three boys and three girls would you most like

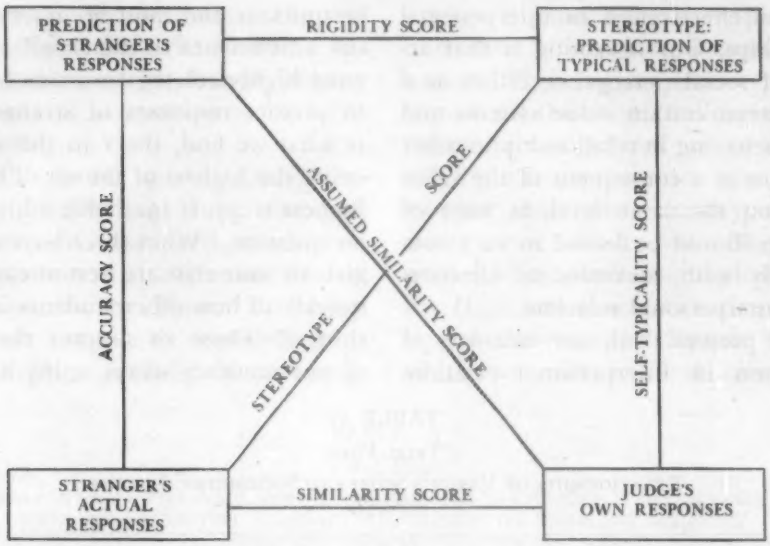


FIG. 1. Kinds of Scores Obtained by Comparing Various Sets of Answers.

to do your school work?" as another measure of effectiveness in interpersonal relationships. The *r*'s of the prediction score with number of mentions in answer to these two questions are .31 and .35, respectively. It seems reasonable to conclude that these *r*'s furnish positive support for the hypothesis that accuracy of social perception correlates positively with effectiveness in interpersonal relationships.

So far we have looked only at the accuracy scores as measures of accuracy in social perception. In Trial Three, however, we raised questions as to whether the accuracy scores indicate (*a*) ability to observe and integrate cues from expressive behavior or (*b*) a kind of persistent expression of familiarity and indoctrination with cultural patterns. Concern with the latter possibility leads us to inspect the correlations of other possible scores with the sociometric indices.

These scores are obtained from the six possible pairings of the four sets of responses obtained in this trial. They are shown schematically in Figure 1 and may be characterized as follows:

1. *Accuracy score.* Obtained by comparing the stranger and the prediction answers, this score is the basic measure of ability to judge strangers with which we have thus far been concerned.

2. *Stereotype score.* Obtained by comparing the stranger and the stereotype answers, this score indicates the appropriateness to the strangers of the judges' stereotypes concerning the interests of male and female seniors at the University High School. The judges' male stereotype answers were scored against the three male strangers' answers; the judges' female stereotype answers were scored against the answers of the three female strangers.

3. *Similarity score.* Obtained by comparing the stranger and the judge answers, this score indicates the similarity between the interests of the judge and of the strangers whose interests he is predicting.

4. *Rigidity score.* Obtained by comparing the prediction and the stereotype answers, this score indicates the degree

to which the judge was "rigid" in the sense of making his predictions for individual strangers, whose expressive behavior he had observed, accord with his stereotype concerning the interests of the cultural subgroup (male or female seniors at the University High School).

5. *Assumed similarity score.* Obtained by comparing the prediction and the judge answers, this score indicates the degree to which the judge ascribed to the strangers the interests he expressed for himself. In some of our analyses, this assumed similarity score will be corrected for the actual similarity between judges' and strangers' interests by subtracting the "similarity" score.

6. *Self-typicality score.* Obtained by comparing the stereotype and judge answers, this score indicates the degree to which the interests expressed by the judge for himself are similar to those which he ascribed to the typical male (for male judges) or female (for female judges) University High School senior.

TABLE 18

TRIAL FOUR

GENERALITY OVER STRANGERS OF VARIOUS SCORES

$N = 40$

| Score | $r_{(A+B+C)(D+E+F)}$ | $r_{corrected}$ |
|---|---|---|
| Stereotype | .52 | .68 |
| Similarity | .44 | .61 |
| Rigidity | .71 | .83 |
| Assumed Similarity | .69 | .82 |

Table 18 shows the $r$'s obtained when these scores based on the first three strangers were correlated with those based on the second three. The $r$'s based on sums of six answer sheets range from .61 to .83 for the corrected values, indicating substantial generality over strangers.

Taking the scores singly, we note that, for all six sociometric indices, the correlations of the stereotype score with the sociometric indices about equal those of the accuracy score. So we have evidence of the psychological equivalence of stereotype accuracy to our initially hypothesized "ability to predict strangers' interests from observations of their expressive behavior." In its relationship to such outside variables as our sociometric indices, our stereotype score behaves so as to acquire much the same significance for interpersonal relationships as our accuracy score possesses. Although its logical validity is different, empirically it seems to bear upon our sociometric indices in much the same way.

The other four scores do not differ markedly from the accuracy and stereotype scores in their correlations with the sociometric indices. The $r$'s for self-typicality, however, tend to be lower, probably because this score is based on comparisons of only two answer sheets per judge. These scores do not reflect accuracy of social perception; rather they measure the tendencies of the judges (*a*) to "perseverate," that is, make the same responses whether predicting, stereotyping, or describing themselves, and (*b*) to be "conventional," that is, resemble the strangers and show saturation with the common culture. Hence we have some indication that it is such tendencies, rather than accuracy of perception as earlier conceived, which accounts for the relationships of these measures to our sociometric indices.

When we look at various combined scores, this interpretation is strengthened. The $r$'s of the "stereotype plus rigidity" scores with the sociometric indices are all higher than those of the accuracy scores. The same is true of the "similarity plus rigidity" score except in its $r$ with the index for "doing schoolwork." Yet both of these combined scores have

perseveration and conventionality in their logic far more than the kind of social perception which the accuracy score would seem to measure.

We can get a logically purer measure of accuracy in social perception of individual strangers by computing an "accuracy minus stereotype" score. This should tell us how much better the predictions were *after* observing expressive behavior than before. Since for most judges the stereotype score exceeded the accuracy score, "accuracy minus stereotype" actually indicates how little the observation detracted from the stereotype score. In any case, the *r*'s with sociometric indices for this combined score are all approximately zero, i.e., much lower than those of the scores that reflect perseveration and conventionality.

How do the correlations of our scores with the sociometric indices compare with those of intelligence tests? Table 17 shows the *r*'s of the Henmon-Nelson IQ's and ACE Psychological Examination scores with the six sociometric indices. Only for the "doing schoolwork" index are the intelligence scores clearly superior (*r*'s of .44 and .60) to our "perseverative conventionality" scores in accounting for variance in sociometric standing. This finding is, of course, highly plausible in terms of the values elicited by this particular sociometric question. The *r* of the ACE score with "being tactful" is .41; this equals the highest *r*, that of the stereotype score, reached by any of our other scores. In the main, however, our measures of social perception or perseverative conventionality clearly excel the intellective measures as correlates of these sociometric indices.

These findings take on added importance when we examine the correlations of our accuracy and stereotype scores with our two measures of intellectual ability. As shown in Table 19, these hover close to zero, the range being from −.07 to .30. The accuracy and stereotype scores are indeed manifestations of a nonintellectual trait.

TABLE 19

TRIAL FOUR

RELATIONSHIPS OF ACCURACY AND STEREOTYPE SCORES TO INTELLECTIVE MEASURES

| | |
|---|---|
| Accuracy Score vs. ACE Psychological Examination ($N=26$) | −.07 |
| Accuracy Score vs. Henmon-Nelson IQ ($N=24$) | −.03 |
| Stereotype Score vs. ACE Psychological Examination ($N=26$) | .30 |
| Stereotype Score vs. Henmon-Nelson IQ ($N=24$) | .14 |

*Correlations of accuracy scores with interests and adjustment.*—For their worth in exploration of the meaning of the accuracy scores, we give in Table 20 their correlations with scores on the Kuder Preference Record, Form CH, and the Bell Adjustment Inventory. None of the *r*'s is statistically significant.

TABLE 20

TRIAL FOUR

CORRELATIONS OF JUDGES' SCORES ON THE KUDER PREFERENCE RECORD, FORM CH, AND THE BELL ADJUSTMENT INVENTORY WITH THEIR ACCURACY SCORES

| Score | Mean | *SD* | *r* with Accuracy Score |
|---|---|---|---|
| Kuder ($N=29$) | | | |
| Outdoor | 62 | 28 | .03 |
| Mechanical | 52 | 31 | −.15 |
| Computational | 37 | 26 | −.25 |
| Scientific | 51 | 28 | .35 |
| Persuasive | 39 | 28 | −.05 |
| Artistic | 61 | 27 | .14 |
| Literary | 61 | 29 | .15 |
| Musical | 54 | 30 | .19 |
| Social Service | 51 | 30 | −.04 |
| Clerical | 34 | 26 | −.35 |
| Bell ($N=23$) | | | |
| Home Adjustment | 5 | 5 | .10 |
| Health Adjustment | 6 | 5 | .14 |
| Social Adjustment | 11 | 7 | −.25 |
| Emotional Adjustment | 8 | 7 | −.20 |
| Total | 31 | 18 | −.09 |

Even the trend of the $r$'s on the Kuder lacks psychological meaning, the pattern of these $r$'s being totally different from that of Trial Three. The $r$'s with the Bell make some sense in that, since higher Bell scores indicate poorer adjustment, the $r$ of $-.25$ between prediction and social adjustment scores is at least in the direction we should have expected. We cannot, however, consider this result to have more than a slight suggestive value.

## VII. SUMMARY AND CONCLUSIONS

Effectiveness in interpersonal relationships may be hypothesized to depend significantly on accuracy of social perception. This study has attempted to develop and explore an approach to the measurement of accuracy of social perception. The operational definition has been used that such accuracy is success in predicting the responses of others on the Kuder Preference Record. Predictions are compared with actual responses of others, and accuracy is scored in terms of the closeness of one to the other.

We arranged four trials of an exercise in which judges observed individuals in brief samples of expressive behavior and then predicted their responses to Kuder Preference Record items. The judges' predictions were scored for their agreement with the actual responses of the strangers.

1. The first task was to ascertain whether scores so obtained possessed some generality from one set of predictions to the next, when sets are made according to groups of stimuli (items) responded to, or according to persons (strangers). On all four trials, the judges were fairly consistent in their success from one set of items to the next, and in all except the second trial they were also fairly consistent from one set of strangers to the next. (In the second trial, the absence of generality over strangers may have been due to the fact that two of the strangers were not members of the same cultural subgroup, undergraduates in teacher-training curricula, as the other four.) In the fourth trial, their success in predicting strangers' responses correlated .71 with success in predicting the responses of three randomly chosen classmates. The conclusion is drawn that the task of predicting strangers' interests after observations of their expressive behavior elicits a performance that has considerable generality.

2. Having established generality, we analyzed the prediction process in terms of the magnitude of five kinds of scores: namely, accuracy, accidental like-sex accuracy, accidental cross-sex accuracy, stereotype accuracy, and stereotype cross-sex accuracy. From this analysis it appeared that success in predicting strangers' responses involved to a slight extent the use of clues from expressive behavior and more a type of familiarity with a particular subculture. Such behavior is reflected in the superiority of the stereotype accuracy to the individual predictive-accuracy scores, even though the latter benefited from being aimed at a particular individual of a given sex.

3. Regardless of the process by which predictive accuracy is achieved, it appears to have meaningful relationships with sociometric status of various kinds. Number of mentions by classmates was taken as an index of effectiveness in various kinds of interpersonal relationships. Six different scores (Predictive Accuracy, Stereotype Accuracy, Similarity, Rigidity, Assumed Similarity, and Self-Typicality) and various combinations of them were found to correlate

meaningfully (about .30 to .45) with these indices.

4. On the other hand, the accuracy and other scores correlate only slightly if at all with two measures of intelligence, with interests as measured by the Kuder test, and with scores on the Bell Adjustment Inventory.

Taking the investigation as a whole, we find considerable promise in the technique of explicit forecasting of strangers' responses as an approach to social perception. Measures so derived exhibit substantial generality, and both a priori and empirical meaning as an aspect of effectiveness in interpersonal relationships.

The technique merits further investigation through the use of motion pictures of different expressive behaviors of a wider variety of strangers. Different kinds of evidence on which to base predictions and different kinds of behavior-to-be-predicted need study. Variations in the cultural similarity of judges and strangers need to be studied as to their relationship to predictive accuracy. Developmental studies of predictive success should be made to plot the growth curve of social perception. And of most practical importance, the significance of ability to make explicit forecasts, or of stereotype accuracy, needs to be examined for different kinds of interpersonal relationships. How do such traits affect relationships between teachers and pupils, clinicians and clients, supervisors and workers, officers and enlisted men? The present study suggests that problems like these can be fruitfully investigated.

## REFERENCES


1. ALLPORT, G. W. *Personality: a psychological interpretation.* New York: Holt, 1937.
2. BENDER, I. E. *A study in integrations of personality by predictions and matching.* Syracuse, New York: Syracuse Univer. Library, 1935.
3. BENDER, I. E., & HASTORF, A. H. The perception of persons: forecasting another person's responses on three personality scales. *J. abnorm. soc. Psychol.*, 1950, **45**, 556-561.
4. CHOWDHRY, KALMA, & NEWCOMB, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 51-57.
5. DYMOND, ROSALIND F. A scale for the measurement of empathic ability. *J. consult. Psychol.*, 1949, **13**, 127-133.
6. ESTES, S. G. Judging personality from expressive behavior. *J. abnorm. soc. Psychol.*, 1938, **33**, 217-236.
7. FISKE, D. W., FIEDLER, F. E., & ISAACS, K. Rigidity in clinical judgments. Paper read at Midwest. Psychol. Ass., Detroit, May, 1950.
8. GAGE, N. L., & SUCI, G. Social perception and teacher-pupil relationships. *J. educ. Psychol.*, 1951, **42**, 144-152.
9. KELLEY, IDA B., & PERKINS, K. J. *How I teach: Purdue teacher's examination.* Minneapolis, Minn.: Educational Test Bureau, 1942.
10. LEEDS, C. H., & COOK, W. W. The construction and differential value of a scale for determining teacher-pupil attitudes. *J. exp. Educ.*, 1947, **16**, 149-159.
11. SMITH, M. B. A study in backward prediction. Unpublished study, Harvard Univer., 1941.
12. STEINMETZ, H. C. Directive psychotherapy: measuring psychological understanding. *J. clin. Psychol.*, 1945, **1**, 331-335.
13. VERNON, P. E. Some characteristics of the judge of personality. *J. soc. Psychol.*, 1933, **4**, 42-58.
14. WOLF, R., & MURRAY, H. A. An experiment in judging personalities. *J. Psychol.*, 1936, **3**, 345-365.
15. WOLFF, W. *The expression of personality.* New York: Harper, 1943.
16. WOOD, H. G. An analysis of social sensitivity, Unpublished Ph.D. thesis. Yale Univer., 1948.


*Accepted for publication May 7, 1952.*

JAN - 7 1959

Psychological Monographs

General and Applied

# Judging Interests from Expressive Behavior

By

N. L. Gage

*Bureau of Educational Research, College of Education, University of Illinois*

Price, $1.00

Edited by Herbert S. Conrad

Published by The American Psychological Association, Inc.

# Psychological Monographs:
## General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*


*Editor*

HERBERT S. CONRAD

*Federal Security Agency*
*Office of Education*
*Washington 25, D.C.*

*Managing Editor*

LORRAINE BOUTHILET

*Consulting Editors*

DONALD E. BAIER
FRANK A. BEACH
ROBERT G. BERNREUTER
WILLIAM A. BROWNELL
HAROLD E. BURTT
JERRY W. CARTER, JR.
CLYDE H. COOMBS
JOHN G. DARLEY
JOHN F. DASHIELL
EUGENIA HANFMANN
EDNA HEIDBREDER
HAROLD E. JONES
DONALD W. MACKINNON
LORRIN A. RIGGS
CARL R. ROGERS
SAUL ROSENZWEIG
ROSS STAGNER
PERCIVAL M. SYMONDS
JOSEPH TIFFIN
LEDYARD R TUCKER
JOSEPH ZUBIN


MANUSCRIPTS should be sent to the Editor.

Because of lack of space, the *Psychological Monographs* can print only the original or advanced contribution of the author. *Background and bibliographic materials must, in general, be totally excluded* or kept to an irreducible minimum. Statistical tables should be used to present only the most important of the statistical data or evidence.

The first page of the manuscript should contain the title of the paper, the author's name, and his institutional connection (or his city of residence). Acknowledgments should be kept brief, and appear as a *footnote* on the first page. No table of contents need be included. For other directions or suggestions on the preparation of manuscripts, see: CONRAD, H. S. Preparation of manuscripts for publication as monographs. *J. Psychol.*, 1948, 26, 447-459.

CORRESPONDENCE CONCERNING BUSINESS MATTERS (such as author's fees, subscriptions and sales, change of address, etc.) should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C.


COPYRIGHT, 1952, BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.